

*THE ACCIDENTAL BEHAVIORIST: A REVIEW OF THE NEW BEHAVIORISM BY  
JOHN STADDON*

WILLIAM M. BAUM

UNIVERSITY OF CALIFORNIA, DAVIS

Staddon contends that behavior analysis may profit from theory, by which he means theory that includes assumed elemental processes allowing derivation of observed patterns of results. Behavior analysts who consider that our field has matured to the point where we are secure in our descriptive grasp of a range of phenomena will tend to agree. Unfortunately, the book's positive message is almost lost in a morass of distracting criticisms of Skinner and behavior analysis in general. Instead of recognizing the growth and maturation of the field, Staddon writes as if the field has stagnated and blames that stagnation on what he takes to be Skinner's antipathy toward theory. Neither has behavior analysis become stagnated, nor should Skinner be blamed for any shortcomings. Instead of acknowledging Skinner's foundational contributions to the field, however, Staddon devotes most of this book to bashing Skinner and fails to distinguish his own view of theory based on internal states from theory in cognitive psychology, treating theoretical models and unseen processes as the "true" subject of inquiry and behavior as only an indicator.

*Key words:* Skinner, behaviorism, theory, internal states, Staddon

This book is about theory. By structure and substance, it makes an argument in favor of theory in behavior analysis, a view that has been controversial for over 50 years. Whether one agrees with the argument or not, the book raises old questions: What is a theory, and what makes a satisfactory theory?

Probably every science began with some sort of folk version. No doubt behavior analysis emerged from folk psychology. The key terms of folk psychology are what philosophers call *intentional*: terms such as "know," "want," and "believe" that seem to imply an agent who knows, wants, and believes. If we say that Tom takes the uptown bus because he wants to go home and believes that bus will take him there, then it is an explanation of sorts. One might say it is based on a theory: that when an agent wants something and believes a course of action will achieve that goal, then the agent will adopt that course of action. Why do we object to such a theory?

The reason is easy to come by. It assigns desire and belief the status of causes, and we have no idea how a desire to go home or a belief in a bus could cause Tom to take the bus. Gilbert Ryle (1949), in his book *The The-*

*ory of Mind*, argued that, from a logical point of view, intentional terms such as know, want, and believe are not causative but summative. Part of Tom's belief in the uptown bus is that he takes it. Part of his wanting to go home is that he goes home. Another part of his belief would be Tom's talking about the uptown bus as a means to get home. Instead of his taking the bus *and* talking about that action constituting evidence of his inner belief, both the action and its description *are* his belief. Ryle criticized the idea that inner belief causes behavior as an example of what he called the paramechanical hypothesis: the idea that mind mechanically causes behavior. He traced the paramechanical hypothesis to the notion that the inner mind animates the external body, which he mockingly referred to as the notion of the "ghost in the machine."

Skinner (1950/1968) also famously attacked the making of theories that invoke inner causes to explain behavior. In an oft-quoted passage, he set out to criticize theories of a certain sort:

... any explanation of an observed fact which appeals to events taking place somewhere else, at some other level of observation, described in different terms, and measured, if at all, in different dimensions. (Skinner, 1950/1968, p. 4)

With this definition, he rejected theories appealing both to mental causes and also to observed or supposed events in the nervous system—in other words, theories that were

---

*The new behaviorism: Mind, mechanism, and society*, by John Staddon. Published in 2001 by Taylor and Francis: 211 pages, with references, author index, and subject index.

Correspondence should be addressed to William M. Baum, 611 Mason #504, San Francisco, California 94108 (e-mail: wmbaum@ucdavis.edu).

“neural, mental, or conceptual” (p. 5). From the context, one gathers that Skinner objected to theories that depend on events in the mind or nervous system (“somewhere else”), observed by introspection or electrodes in the brain (“at some other level”), described in terms of ideas, processes, or synapses (“different terms”), and either never measured (mental causes) or measured in dimensions having nothing to do with behavior (e.g., membrane potentials).

In historical perspective, sciences appear to adopt different standards about theories at different stages in their development. Skinner’s critique of learning theories arose because he saw that traditional theories about behavior had outlived their usefulness in the light of new methods of generating data and of new types of data being generated. The possibilities for studying operant behavior recorded with counters and cumulative records seemed virtually endless. They totally eclipsed the old runways and mazes. With all these new data, Skinner was saying, we have no need of the old scaffolding, because we can develop new theories that consist of “a formal representation of the data reduced to a minimal number of terms” (p. 21). As he never actually offered such a theory himself, his suggestion remains open to interpretation, and 50-odd years later, we may wonder whether we have developed such theories and whether we might be ready for some other types.

No doubt operant methods have generated a great deal of data. One may easily set up experiments and drown in data. Since Skinner, perhaps as a defense against drowning, we have developed formal representations that help us to ask more pointed questions. The book *Schedules of Reinforcement*, by Ferster and Skinner (1957), illustrates a certain type of exploratory research, based on asking the question, “What if I do this?” over and over. It contains an enormous quantity of information, but hardly organized at all, except by the schedules tried. Nowadays, we have a more quantitative analysis, stemming mostly from Herrnstein’s (1961) discovery of the matching law. Indeed, the matching law, in its various forms (Baum, 1974, 1979; Herrnstein, 1970, 1974), would seem to be exactly the sort of formal representation of data that Skinner advocated.

Oddly enough, Skinner (1986) objected to the matching law. Perhaps the reason was that he actually had a theory, but of a different sort than a formal representation of data. He had an idea of a mechanism that would generate all those beautiful cumulative records. In his famous paper on “Superstition in the Pigeon,” and in the book with Ferster, he adopted the theory that delivery of a reinforcer reinforces whatever behavior immediately precedes it and proposed that this mechanism would account for all changes in behavior (Ferster & Skinner, 1957; Skinner, 1948/1968). Researchers who study the matching law are similarly dissatisfied with only formal representation; they want to know what underlying mechanism might account for the observed regularity.

Hence we may infer that another type of theory exists besides formal representation: theories that allow one to derive the formal representation. Although one could argue that such a theory is itself a formal representation, for our discussion of what a theory is and what is a satisfactory theory, we may usefully distinguish between a purely descriptive summary and a theory that allows one to derive that descriptive summary from more elemental processes. Such a theory still may escape Skinner’s strictures, because the elemental processes (reinforcer-response contiguity for Skinner, and, for example, reinforcer tracking for the matching law [Davison & Baum, 2000]) may take place in, and be measured in terms of, behavior.

So, what about Staddon’s view—theoretical behaviorism? It amounts to the idea that theories of the elemental sort (that allow derivation of data) are permissible in behavior analysis. In particular, Staddon advocates dynamic, real-time theories. He presents as an example a model of habituation employing the notion of cascaded leaky integrators that he has presented a few times before. It is a good example because it allows prediction of a seemingly paradoxical result: that even though high rates of stimulation produce greater habituation, the habituation produced by lower rates of stimulation dissipates more slowly. I can easily imagine the satisfaction the accomplishment must have brought.

To read Staddon’s book, however, one would think that no other behavior analyst looks with favor on such theories. He refers

to almost no other behavior analyst except himself and his students. He writes as if he alone sees how to bring theory to the science and as if all behaviorists since Hull and Tolman have labored in the darkness of an irrational and implacable hostility to theory. Here is Staddon to save the day. But never mind; the important question is how well he makes the case.

The weaknesses of the book stem from Staddon's structuring it in light of "old wrong, new right." He overlooks the development of the field. Research in behavior analysis has grown quantitative, particularly in areas such as choice, detection, timing, and behavioral economics. These developments have taken us to the point where theory makes sense—is no longer a distraction, a snare, and a delusion. Instead of considering that Skinner's strictures about theory may have been constructive in 1950 and that the field has developed since to the point where it is ready for theory, Staddon takes to bashing Skinner, lending the book a sadly negative tone. For example, the quotation about "events taking place somewhere else," like most of Skinner's writing is open to multiple interpretations, but Staddon chooses to interpret it in only one way: as forbidding virtually any type of theory. In historical context, Skinner's statement may be seen to refer to premature modeling in the absence of data, the sort of theories that learning psychologists were spinning with scant support from data. Staddon, however, uses his special interpretation to call Skinner's warning "antitheoretical," "bizarre," and "theoretical seppuku."

The best chapters are the last three: one on cognitive psychology, one on theory, and one on consciousness. Any behaviorist will enjoy the chapter on cognitive psychology, because it assembles criticisms from philosophy and artificial intelligence. The presentation of theoretical behaviorism, the next-to-last chapter, is difficult to interpret, because Staddon introduces the idea of internal states but never clarifies their ontological status. First he portrays hypothetical constructs as "place holders" in quantitative models. That sounds all right; state variables are useful in models. He goes on, however, to call them "internal states" and to suggest that models tell how the brain behaves and that the physiologists' job is to "figure out how the brain can behave

in the way our model says it does" (Staddon, 2001a, p. 153). Despite his claims to the contrary, Staddon, like the cognitivists, thinks he is studying states and mechanisms in the brain.

As the standard for deciding among models, Staddon shies away from prediction, and appeals instead to parsimony. Parsimony, however, like aesthetics, has its own problems; one person's parsimony is another person's oversimplification. Perhaps aware of this, Staddon waffles and winds up equating models with formal representations of data, the view of which Skinner, inspired by Mach's (1883/1960) *Science of Mechanics*, also approved. Staddon comments disingenuously, "a formulation not far from TB" (theoretical behaviorism; Staddon, 2001a, p. 154). More like indistinguishable—but that would raise the question of what all the fuss is about.

As I was reading the book, that question occurred to me many times. The most bizarre and troubling aspect of it is that, following an initial chapter praising Hull and Tolman, fully four chapters—half the book—are devoted to assessing and mostly attacking Skinner. Instead of presenting his own views in a positive way, Staddon goes on page after page excoriating Skinner. I found myself objecting, even though I agree with most of the criticisms. I agree that Skinner indulged in premature extrapolation, that he was wrong about the facts on punishment, that he made unlikely, possibly irresponsible, recommendations. Staddon's critique of Skinner's (1948/1968) superstition paper as a piece more polemical than scientific is excellent, except for the tone. There and throughout, his choice of words is sarcastic and even snide. When he calls Skinner a rhetorical genius, it comes out as an insult.

Staddon's attack, like his portrayal of Skinner's view of theory, is often distorted and is much too harsh. He misrepresents Skinner's views on responsibility, for example, promoting himself as wiser or more humane. Skinner's objections were to mentalism and autonomous man, both of which Staddon also would reject, but in a discussion parallel to the one in my own book, *Understanding Behaviorism* (Baum, 1994), Staddon redefines responsibility in practical terms, with the difference that he then uses his redefinition to argue illogically that Skinner was wrong. Like

Skinner, he argues, too, that freedom is a practical affair, mentioning Skinner's stress on feeling free, but calling it subjective, and then claiming that Skinner was in some (unexplained) way wrong about that, too.

On top of criticism that, if misguided, is at least grounded, Staddon goes on to ridiculous extremes. He includes a long discussion of the necessity and virtues of punishment in social policy that rests entirely on the unexamined assumption that punishment works as a deterrent. I was reminded of an article (Staddon, 1995) that he published in the *Atlantic Monthly* magazine some years ago, in which he presented similar arguments in favor of the death penalty, completely ignoring that all research so far indicates it is ineffective as a deterrent. (Perhaps he should have followed his own advice about premature extrapolation.) The capper, however, was his treatment of Barbara Herrnstein Smith, a postmodern social constructionist well deserving of criticism. I enjoyed Staddon's criticism of postmodernism and Smith, even though it seemed like going off on a tangent. He points out the illogic of the central premise that there is no such thing as truth. (Is it true or false?) He responds to the accusation that scientific theories are often judged by other than objective standards. (So what?) I was distressed, however, that he tries to blame Smith's excesses on Skinner. Worse, he does this by innuendo and conjunction. First, we read that Skinner's ideas about truth "have taken root in some strange lands, most recently in the politically correct thickets of literary theory" (Staddon, 2001a, p. 67). Fancy prose, this, but are we supposed to believe that Skinner was responsible for political correctness? Then he tells us that Skinner is "cited with approval by more than one fan of the 'postmodern aesthetic,' " as if being cited were a sin. Staddon goes on to say that the views of "Skinner and Skinnerians" (note the insertion of fellow travelers) resemble those of "postmodernists/deconstructionists and relativists such as Foucault, Derrida, and Latour" in that all "are skeptical of rationalism, objectivity, and the idea of an independent, external reality" (Staddon, 2001a, p. 80). As someone who has searched long and hard for evidence of Skinner's views about external reality, I would challenge anyone to find support in his writings even for skepticism, let

alone rejection. Staddon is simply trying to tar Skinner with the postmodernist brush so that he can justify including a critique of relativism that would otherwise seem out of place. He calls Barbara Herrnstein Smith "Skinner-influenced." She knew Skinner, but that hardly constitutes grounds for blaming her excesses on Skinner, and Staddon offers no evidence.

As I read through his diatribe against Skinner and Skinnerians, I wondered why Staddon would write this. What purpose could it serve? Most likely, it is a political move; he wants to distance himself from Skinner and to curry favor with the anti-Skinner psychologists and philosophers. This fails, however, because those folks rarely read or understand Skinner's writings, and Staddon winds up simply strengthening their prejudices. The harder but more honorable way would have been to take from Skinner what is valuable while acknowledging the source.

Skinner, in fact, wrote much that was valuable for a science of behavior. In what appears as an anomalous page in which he acknowledges Skinner's contributions, Staddon calls him a "brilliant experimenter" and mentions that "Skinner provided a conceptual framework for understanding learning that (I believe) has yet to be fully explored," although he can't help adding, "even though his strictures against theory prevented him from exploiting it himself and impeded the efforts of others to do so" (Staddon, 2001a, p. 122). This is damning with faint praise. Skinner invented the "Skinner box," but his experiments were much less significant than what he found to say about them. The science we practice today would have been impossible without Skinner's invention of the concepts of operant and stimulus control and his emphasis on rate as a dependent variable. Staddon himself owes a huge intellectual debt to Skinner, if he would but consider. If Staddon has seen far (and that is an open question), it is because he stood on the shoulders of giants, one of whom was Skinner and another a student of Skinner's.

In the included preface to the first edition, which was a shorter and much less anti-Skinner book, Staddon writes, "I never used to think of myself as a behaviorist, but now I see that I have been ignoring the evidence" (Staddon, 2001a, p. xv). On the occasion of

writing, he discovered himself to be a behaviorist. Would that he had embraced this realization. Instead, he resists it and writes as an outsider, referring to behaviorists as “they” and disparaging their institutions. Their journals, he reports, publish research “in the Skinnerian tradition” (Staddon, 2001a, p. 122). From the context, we are left in no doubt that this includes the *Journal of the Experimental Analysis of Behavior*, in which Staddon himself has often published. The implication appears to be that behavior analysts (presumably excepting Staddon) have made no conceptual advances beyond Skinner’s framework. This is self-aggrandizing and false. About research on choice since Herrnstein’s (1961) discovery of the matching law, Staddon comments, “. . . a whole industry arose to study the topic of choice” (p. 39), thus to denigrate and dismiss decades of thoughtful experimentation as plodding and directionless (the implication of “industry”), presumably because he thinks it made no theoretical advance. Perhaps if Staddon made more effort to notice the research of other behavior analysts, he would be less inclined to present it so falsely.

The positive contribution of the book lies in a section of chapter 6 (“Mind and Mechanism”) on theoretical behaviorism and chapters 7 (“Internal States”) and 8 (“Consciousness and Theoretical Behaviorism”). The presentation of theoretical behaviorism in chapters 6 and 7 takes up only 19 pages (less than 10% of the book). In it, Staddon makes the transition from state variables to internal states without justification and leaves the reader in confusion as to exactly what he intends about their status. In chapter 6, he attacks cognitivism on two grounds: philosopher John Searle’s (1992) criticism of the brain-computer analogy, and research on artificial intelligence that demonstrates the emergent intelligence of many unintelligent units operating in concert, implying the lack of need for internal or mental representations. I welcomed the criticisms of the computer analogy and representations, but noticed that Staddon carefully avoided criticizing the antibehavioral aspect of cognitivism: The denial of behavior as a subject matter in preference to an idea that behavior is only the evidence of inner processes that are the real subject matter. The reason is easy

to find: Staddon embraces that very idea. He tells us first that theoretical models incorporate internal states and second that “These models *are* the behavior . . . what the organism is ‘doing,’ described in the most colorless, direct way possible” (Staddon, 2001a, p. 144). To me, this statement seems indistinguishable from the cognitivists’ competence-performance distinction. How would deviations of observed behavior from the model’s predictions be interpreted? After that, Staddon brings out the example that appears in previous publications: cascaded leaky integrators as a model of habituation. He shows how a model composed of two cascaded integrators predicts rate dependence. Curiously, however, he fits the model to no data, and the only data shown are grouped—the proportion of a group of nematodes responding to a tap on the container—whereas the model predicts strength of response in an individual organism. This is suggestive, but hardly compelling.

The final chapter on consciousness is the best, although only loosely connected to the rest of the book. Staddon explains the “Turing-test view of consciousness: If a subject can make the appropriate verbal report, we will grant him consciousness” (Staddon, 2001a, pp. 161–162). He goes on to say, “Theoretical behaviorism accepts the Turing-test view of consciousness: If you act conscious (as assessed by an admittedly fallible human inquisitor), you *are* conscious” (p. 162). Unfortunately, he never explains why and never acknowledges the significance of “admittedly fallible.” Never mind, I liked it anyway. Staddon leaves off his axe grinding and discusses consciousness in a way consistent with other contemporary behavioristic treatments (e.g., Baum, 1994; Rachlin, 1994). He presents and uses a three-part distinction among the domains of experience (the subjective, about which science has nothing to say), physiology (the functioning of the brain, about which much is still to be learned), and behavioral data (intersubjectively verifiable reports). Armed with this, he debunks several antibehavioral and indeed antiscientific arguments based on consciousness as a thing and as a cause.

I am sorry I could find so little positive to say about this book. I found its negative tone offensive, its presentation of the place of the-



ory unclear and ambiguous, and much of the philosophical discussion imprecise and half-baked. This was true despite my agreement with most of the essential points. In particular, I agree that behavior analysis has developed to the point where it is ready for theory, as Skinner suggested it would. Maturity in a science may be gauged by the degree to which its data are organized by the application of mathematics. Models like Staddon's allow the development of mathematical treatments of data. That is their value. Who reifies the variables in a mathematical model by calling them internal states makes a mistake. I notice that Staddon has another book on theory (2001b), *Adaptive Dynamics: The Theoretical Analysis of Behavior*. I look forward to seeing the theories without the polemics.

## REFERENCES

- Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, 22, 231–242.
- Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior*, 32, 269–281.
- Baum, W. M. (1994). *Understanding Behaviorism: Science, Behavior, and Culture*. New York: HarperCollins.
- Davison, M., & Baum, W. M. (2000). Choice in a variable environment: Every reinforcer counts. *Journal of the Experimental Analysis of Behavior*, 74, 1–24.
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.
- Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4, 267–272.
- Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13, 243–266.
- Herrnstein, R. J. (1974). Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, 21, 159–164.
- Mach, E. (1960). *The science of mechanics: A critical and historical account of its development*. LaSalle, IL: Open Court. (Translation of the ninth German edition, 1933. Original work published 1883.)
- Rachlin, H. (1994). *Behavior and mind: The roots of modern psychology*. New York: Oxford University Press.
- Ryle, G. (1949). *The concept of mind*. Chicago: University of Chicago Press.
- Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT/Bradford.
- Skinner, B. F. (1948). "Superstition" in the pigeon. *Journal of Experimental Psychology*, 38, 168–172. Reprinted in Catania, A. C. (1968). *Contemporary research in operant behavior* (pp. 62–64). Glenview, IL: Scott, Foresman and Company.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, 57, 193–216. Reprinted in Catania, A. C. (1968) *Contemporary research in operant behavior* (pp. 4–21). Glenview, IL: Scott, Foresman and Company.
- Skinner, B. F. (1986). Some thoughts about the future. *Journal of the Experimental Analysis of Behavior*, 45, 229–235.
- Staddon, J. (1995). On responsibility and punishment. *Atlantic Monthly*, 275(2), 88–94.
- Staddon, J. E. R. (2001a). *The new behaviorism*. Philadelphia: Taylor & Francis.
- Staddon, J. E. R. (2001b). *Adaptive dynamics*. Cambridge, MA: MIT/Bradford.